

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-202318
 (43)Date of publication of application : 27.07.2001

(51)Int.Cl. G06F 13/00
 G06F 12/00
 G06F 15/177
 H04L 12/46
 H04L 12/28
 H04L 12/40

(21)Application number : 2000-014337

(71)Applicant : HITACHI KOKUSAI ELECTRIC INC

(22)Date of filing : 24.01.2000

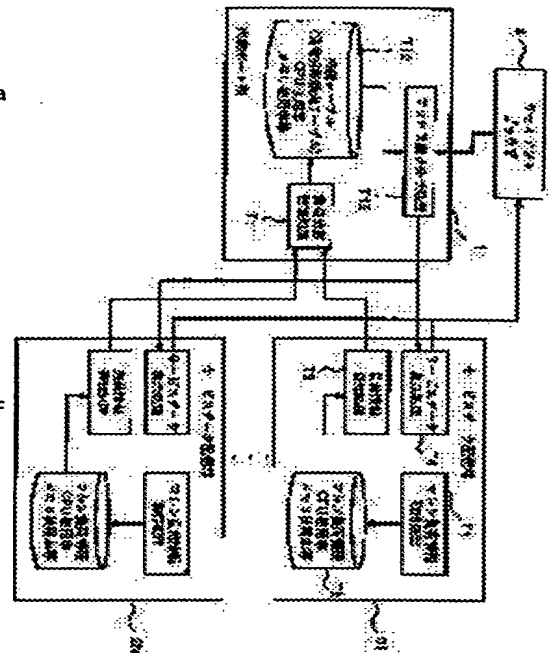
(72)Inventor : TATEDA HIROAKI

(54) DATA DISTRIBUTION SYSTEM

(57)Abstract:

PROBLEM TO BE SOLVED: To automatically distribute the load of a server device while using the exact load of each of server devices in a data distribution system for transmitting data through a network to a client device by a server device selected out of plural server devices by a managing device.

SOLUTION: Each of server devices S1-SN is provided with a transmitting means for transmitting data to a client device 4 by transmitting processing performed by a CPU while using an internal memory and a reporting means reports the activity ratio of the CPU and the activity ratio of the memory to a managing device 1 at prescribed timing. In the managing device, a managing means manages the activity ratio of the CPU and the activity ratio of the internal memory reported from each of server devices and corresponding to a request transmitted from the client device through the network, a transmission control means transmits data to the client device by the server device of small load determined while using the activity ratio of the CPU and the activity ratio of the internal memory.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.**** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

PRIOR ART

[0002] [Description of the Prior Art] For example, in the data distribution system of the Internet, the client equipment by the side of a user and the server equipment by the side of service data supply are formed, and transmitting data (service data) to client equipment through the Internet from server equipment is performed according to the demand transmitted through the Internet from client equipment.

[0003] By the way, since server equipment was formed only for one set, for example, when access to server equipment failed in the conventional data distribution system, the user side needed to perform the retry of access etc. and needed to publish the display demand of HTML data etc. again in it. For this reason, when the load access to server equipment carries out congestion etc., and starts the server equipment concerned, for example was expensive, there was a problem that the time-out of access processing will occur frequently by the user side.

[0004] The data distribution system which formed the mirror server equipment which stored the same data for example, as server equipment as what avoids such a problem is known. In this data distribution system, when a user side changes an access place to the lighter one of a load in server equipment and mirror server equipment, a user side can be accessed to equipment with a lighter load (server equipment or mirror server equipment), can receive data from the equipment concerned, and, thereby, can distribute a load with server equipment and mirror server equipment.

[0005] In addition, an example of the data distribution system of the Internet which connected and formed the above server (Maine server) equipments 17 and mirror server equipment 18 in drawing 5 through Ethernet 16 is shown. In this data distribution system, client equipment 11 is connected with the Internet 14 through the

modem 12 grade and the provider 13 grade, and server equipment 17 and mirror server equipment 18 are connected with the Internet 14 through Ethernet 16 and a router (or proxy server) 15.

[0006] However, in the data distribution system which formed mirror server equipment as mentioned above, the user side had the troublesomeness that the address of an access place had to be changed according to the equipment (server equipment or mirror server equipment) with which the load profile initiation of each equipment cannot be grasped unless it accesses to each equipment (server equipment or mirror server equipment) actually, and a user side wishes to access.

[0007] As what cancels such troublesomeness, in the server system for the Internet indicated by JP,10-27148,A (henceforth reference 1), a load level is set up according to the response time of the monitor command which a management server transmits periodically to a server, and it performs choosing a server with the lightest load.

[0008] Moreover, in the dynamic restructuring of the Network Server indicated by JP,10-105500,A (henceforth reference 2), for example, in the client/server system which prepared two or more servers, when each server is maintaining the processing load in the form of the count of access (demand) served for per unit time amount and the processing load of a certain server becomes superfluous, it performs that the server concerned redirects a client demand to other servers.

[0009] Moreover, although the load of each server is not supervised, network load balancing for the multi-computer server indicated by JP,11-143804,A, for example is introduced. In network load balancing for this multi-computer server, it performs determining the server to which a message dispatcher should answer the parameter (for example, the accumulation count of a packet size and/or the number of average packets per second) showing a network load, and should dispatch the demand from a client out of two or more servers.

[Translation done.]

* NOTICES *

Japan Patent Office is not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

MEANS

[0013] [Means for Solving the Problem] In order to attain the above-mentioned object, data is made to transmit to client equipment through a network in a data distribution system concerning this invention according to a demand transmitted to management equipment through a network from client equipment as follows with the server equipment which management equipment chose from two or more server equipments.

[0014] Namely, CPU is equipped with a transmitting means to transmit data to client equipment through a network by transmitting processing performed using an internal memory, with each server equipment. While an advice means notifies an activity ratio of CPU, and an activity ratio of an internal memory to management equipment to predetermined timing, with management equipment An activity ratio of CPU it is notified from each server equipment that a management tool is, and an activity ratio of an internal memory are managed. Data is made to transmit to client equipment with server equipment with a small load with which a transmission-control means is determined using an activity ratio of CPU, and an activity ratio of an internal memory according to a demand transmitted from client equipment.

[0015] Therefore, an activity ratio of CPU with which each server equipment was equipped, and an activity ratio of an internal memory are managed, and since server equipment which performs data transmission based on a load determined using these is chosen, a load of server equipment can be distributed based on a actual load of each server equipment, and an exact load in agreement. In addition, in this invention, not only a data distribution system such but management equipment equipped with the above management tools or a transmission-control means is offered.

【特許請求の範囲】

【請求項 1】 クライアント装置からネットワークを介して管理装置へ送信される要求に応じて、管理装置が複数のサーバ装置の中から選択したサーバ装置によりデータをネットワークを介してクライアント装置へ送信させるデータ配信システムであって、各サーバ装置には、CPU が内部メモリを用いて行う送信処理によりデータをネットワークを介してクライアント装置へ送信する送信手段と、CPU の使用率及び内部メモリの使用率を所定のタイミングで管理装置に通知する通知手段と、を備え、管理装置には、各サーバ装置から通知される CPU の使用率及び内部メモリの使用率を管理する管理手段と、クライアント装置から送信される要求に応じて、CPU の使用率及び内部メモリの使用率を用いて決定される負荷が小さいサーバ装置によりデータをクライアント装置へ送信させる送信制御手段と、を備えたことを特徴とするデータ配信システム。

【請求項 2】 クライアント装置からネットワークを介して送信される要求に応じて、複数のサーバ装置の中から選択したサーバ装置の CPU が内部メモリを用いて行う送信処理によりデータをネットワークを介してクライアント装置へ送信させる管理装置であって、各サーバ装置から所定のタイミングで通知される CPU の使用率及び内部メモリの使用率を管理する管理手段と、クライアント装置から送信される要求に応じて、CPU の使用率及び内部メモリの使用率を用いて決定される負荷が小さいサーバ装置によりデータをクライアント装置へ送信させる送信制御手段と、を備えたことを特徴とする管理装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、管理装置が複数のサーバ装置の中から選択したサーバ装置によりデータをネットワークを介してクライアント装置へ送信させるデータ配信システムやこのような管理装置に関し、特に、各サーバ装置の CPU の使用率及び内部メモリの使用率を用いてサーバ装置の負荷を自動的に分散させる技術に関する。

【0002】

【従来の技術】 例えばインターネットのデータ配信システムでは、ユーザ側のクライアント装置と、サービスデータ供給側のサーバ装置とを設け、クライアント装置からインターネットを介して送信される要求に応じて、サーバ装置からデータ（サービスデータ）をインターネットを介してクライアント装置へ送信することが行われる。

【0003】 ところで、従来のデータ配信システムではサーバ装置が 1 台しか設けられていなかったため、例え

ばサーバ装置へのアクセスが失敗した場合には、ユーザ側がアクセスのリトライ等を行って HTML データ等の表示要求を再度発行する必要があった。このため、例えばサーバ装置へのアクセスが混雑等して当該サーバ装置にかかる負荷が高いときには、ユーザ側でアクセス処理のタイムアウトが頻発してしまうといった問題があった。

【0004】 このような問題を回避するものとして、例えばサーバ装置と同一のデータを格納したミラーサーバ装置を設けたデータ配信システムが知られている。このデータ配信システムでは、サーバ装置とミラーサーバ装置との中で負荷の軽い方へユーザ側がアクセス先を切り替えることにより、ユーザ側は負荷の軽い方の装置（サーバ装置或いはミラーサーバ装置）へアクセスして当該装置からデータを受信することができ、これにより、サーバ装置とミラーサーバ装置とで負荷を分散させることができる。

【0005】 なお、図 5 には、上記のようなサーバ（メインサーバ）装置 17 及びミラーサーバ装置 18 をイーサネット 16 を介して接続して設けたインターネットのデータ配信システムの一例を示してある。このデータ配信システムでは、クライアント装置 11 がモデム 12 等及びプロバイダ 13 等を介してインターネット 14 と接続されており、また、サーバ装置 17 及びミラーサーバ装置 18 がイーサネット 16 及びルータ（又は、プロキシサーバ）15 を介してインターネット 14 と接続されている。

【0006】 しかしながら、上記のようにミラーサーバ装置を設けたデータ配信システムでは、ユーザ側は実際に各装置（サーバ装置或いはミラーサーバ装置）へアクセスしてみないと各装置の負荷状況を把握することができず、また、ユーザ側がアクセスを希望する装置（サーバ装置或いはミラーサーバ装置）に応じてアクセス先のアドレスを変更しなければならないといった煩わしさがあった。

【0007】 このような煩わしさを解消するものとして、例えば特開平 10-27148 号公報（以下で、文献 1 とする）に記載されたインターネット用サーバシステムでは、管理サーバがサーバに対して定期的に送信する監視コマンドの応答時間に応じて負荷レベルを設定し、最も負荷の軽いサーバを選択することを行う。

【0008】 また、例えば特開平 10-105500 号公報（以下で、文献 2 とする）に記載されたネットワーク・サーバの動的再構成では、複数のサーバを設けたクライアントサーバシステムにおいて、各サーバは単位時間当たりにサービスされたアクセス（要求）の回数の形で処理負荷を維持しており、或るサーバの処理負荷が過剰になった場合に当該サーバがクライアント要求を他のサーバに対してリダイレクトすることを行う。

【0009】 また、各サーバの負荷を監視するものでは

ないが、例えば特開平11-143804号公報に記載されたマルチ・コンピュータ・サーバのためのネットワーク負荷均衡化を紹介しておく。このマルチ・コンピュータ・サーバのためのネットワーク負荷均衡化では、メッセージ・ディスパッチャがネットワーク負荷を表すパラメータ（例えばパケット長の累積カウントおよび／または1秒当たりの平均パケット数）にตอบสนองして、クライアントからの要求をディスパッチすべきサーバを複数のサーバの中から決定することを行う。

【0010】

【発明が解決しようとする課題】しかしながら、上記のような従来のデータ配信システムでは、複数のサーバ装置を管理する管理装置が各サーバ装置の実際の負荷を正確に検出することができないため、サーバ装置の負荷分散の効果が十分に得られないといった不具合があった。

【0011】例えば上記文献1に記載されたインターネット用サーバシステムのように応答時間から負荷レベルを設定する仕方では、一例として管理サーバとサーバとを接続するLAN（Local Area Network）の混み具合等によっても応答時間が変化してしまうといった問題があり、各サーバの負荷レベルを正確に検出することはできなかった。また、例えば上記文献2に記載されたクライアントサーバシステムのような形で処理負荷を維持する仕方においても、同様に、各サーバに備えられたCPU等の実際の負荷と一致する処理負荷が必ずしも検出されないといった問題があり、また、このシステムでは管理サーバが設けられずにサーバ間でのリダイレクションが行われるため、データ配信処理に遅延が生じ易いといった問題があった。

【0012】本発明は、このような従来の課題を解決するためになされたもので、管理装置が複数のサーバ装置の中から選択したサーバ装置によりデータをネットワークを介してクライアント装置へ送信させるに際して、各サーバ装置のCPU及びメモリの正確な負荷に基づいてサーバ装置の負荷を分散させることができるデータ配信システムや、このような管理装置を提供することを目的とする。

【0013】

【課題を解決するための手段】上記目的を達成するため、本発明に係るデータ配信システムでは、次のようにして、クライアント装置からネットワークを介して管理装置へ送信される要求に応じて、管理装置が複数のサーバ装置の中から選択したサーバ装置によりデータをネットワークを介してクライアント装置へ送信させる。

【0014】すなわち、各サーバ装置では、CPUが内部メモリを用いて行う送信処理によりデータをネットワークを介してクライアント装置へ送信する送信手段を備え、通知手段がCPUの使用率及び内部メモリの使用率を所定のタイミングで管理装置に通知する一方、管理装置では、管理手段が各サーバ装置から通知されるCPU

の使用率及び内部メモリの使用率を管理し、送信制御手段がクライアント装置から送信される要求に応じて、CPUの使用率及び内部メモリの使用率を用いて決定される負荷が小さいサーバ装置によりデータをクライアント装置へ送信させる。

【0015】従って、各サーバ装置に備えられたCPUの使用率及び内部メモリの使用率が管理されて、これらを用いて決定される負荷に基づいてデータ送信を行うサーバ装置が選択されるため、各サーバ装置の実際の負荷と一致する正確な負荷に基づいてサーバ装置の負荷を分散させることができる。なお、本発明では、このようなデータ配信システムばかりでなく、上記のような管理手段や送信制御手段を備えた管理装置も提供する。

【0016】

【発明の実施の形態】本発明に係る一実施例を図面を参照して説明する。なお、本例では、本発明に係るデータ配信システムと本発明に係る管理装置とをまとめて説明する。図1には、本発明に係るデータ配信システムの概略的な構成例を示してあり、このデータ配信システムには、例えばコンピュータ等から構成された代表ポート部1と、同様にコンピュータ等から構成された複数（本例では、N台）のサービスデータ配信部S1～SNと、これら複数のサービスデータ配信部S1～SN同士を接続するとともにこれら複数のサービスデータ配信部S1～SNと代表ポート部1とを接続するLAN2とが設けられている。

【0017】また、同図には、例えばコンピュータ等から構成されたクライアント装置4や、代表ポート部1とクライアント装置4とを接続するインターネット3を示してあり、本例では、複数のサービスデータ配信部S1～SNとクライアント装置4ともLAN2やインターネット3を介して接続されている。なお、同図では、1つのクライアント装置4のみを示したが、本例では、多数のクライアント装置がインターネット3に接続されている。

【0018】ここで、代表ポート部1は本発明に言う管理装置の一実施例に相当し、サービスデータ配信部S1～SNは本発明に言うサーバ装置の一実施例に相当し、クライアント装置4は本発明に言うクライアント装置の一実施例に相当する。また、代表ポート部1や各サービスデータ配信部S1～SNやクライアント装置4では、それぞれに備えられたCPUが内部メモリを用いて各種のアプリケーションプログラムを起動して実行することにより、当該プログラムに対応した各種の処理を実現する。

【0019】まず、上記したクライアント装置4や上記した代表ポート部1や上記した各サービスデータ配信部S1～SNに備えられた機能を概略的に説明する。すなわち、クライアント装置4は、例えばインターネット3上のデータを閲覧するためのソフトウェアであるブラウ

ザを備えており、インターネット3を介して代表ポート部1へアクセスして、自己の表示画面上への表示を希望するデータを要求する信号（表示要求）を代表ポート部1へ送信する機能を有している。また、クライアント装置4は、サービスデータ配信部S1～SNからLAN2やインターネット3を介して送信される表示データを受信し、当該表示データを自己の表示画面上に表示出力する機能を有している。なお、表示要求は、例えばクライアント装置4を操作するユーザからの指示により発信させられる。

【0020】また、代表ポート部1は、クライアント装置4からのアクセスを受け付ける窓口となっており、各サービスデータ配信部S1～SNから通知されるマシン負荷情報に基づいて各サービスデータ配信部S1～SNの負荷を監視する機能を有しているとともに、監視している各サービスデータ配信部S1～SNの負荷状況に基づいて負荷が小さいサービスデータ配信部S1～SNを選択することにより、クライアント装置4からのアクセス（表示要求）を各サービスデータ配信部S1～SNに振り分ける機能を有している。

【0021】また、各サービスデータ配信部S1～SNは、クライアント装置4からのアクセス（表示要求）に応じて当該クライアント装置4へ供給するための種々なデータ（サービスデータ）を格納した外部メモリを備えており、クライアント装置4からの表示要求に応じた表示データ（例えばHTMLデータ）をLAN2やインターネット3を介して当該クライアント装置4へ送信する機能を有している。ここで、本例では、各サービスデータ配信部S1～SNのいずれもが同一の表示データをクライアント装置4へ供給することができる構成となっている。また、上述したように、本例では、クライアント装置4からの表示要求は代表ポート部1を介して各サービスデータ配信部S1～SNへ伝送される。

【0022】また、各サービスデータ配信部S1～SNは、自己に備えられたCPUの使用率や内部メモリの使用率に関する情報（マシン負荷情報）を例えば所定の期間（一定の時間間隔）毎にLAN2を介して代表ポート部1に通知する機能を有している。ここで、CPUの使用率とは、例えばCPUが有する処理能力の中で実際に利用されている処理能力の割合を示しており、具体的には、例えばCPUが単位時間当たり実行することが可能な最大のプロセス数の内で実際に実行されているプロセス数の割合を示す。また、内部メモリの使用率とは、例えば内部メモリが有する記憶容量の中で実際に利用されている記憶容量の割合を示しており、具体的には、例えば内部メモリが記憶することが可能な最大のデータ量の内で実際に記憶されているデータ量の割合を示す。

【0023】次に、図2を参照して、上記した代表ポート部1や上記した各サービスデータ配信部S1～SNにより行われる負荷分散処理の具体例を示す。なお、同図

には、本例の負荷分散処理に関して、各サービスデータ配信部S1～SNにより行われる処理T1～T4や代表ポート部1により行われる処理T11～T13を概念的に示してある。また、本例では各サービスデータ配信部S1～SNは同様な構成を有して同様な処理を行うため、同図では、1つのサービスデータ配信部S1により行われる処理のみについて符号T1～T4を付してある。

【0024】まず、各サービスデータ配信部S1～SNによりマシン負荷情報を取得して当該マシン負荷情報を代表ポート部1に通知する処理の具体例を示す。すなわち、各サービスデータ配信部S1～SNでは、例えば所定の期間毎に自己に備えられたCPUの使用率や内部メモリの使用率をマシン負荷情報として取得し（処理T1）、当該マシン負荷情報を記憶等して管理するとともに（処理T2）、当該マシン負荷情報をLAN2を介して代表ポート部1へ送信することで当該マシン負荷情報を代表ポート部1に通知する（処理T3）。

【0025】次に、代表ポート部1により各サービスデータ配信部S1～SNのマシン負荷情報を管理して、クライアント装置4からの表示要求を各サービスデータ配信部S1～SNに振り分ける処理の具体例を示す。すなわち、代表ポート部1では、各サービスデータ配信部S1～SNから送信されるマシン負荷情報を受信し、受信した各サービスデータ配信部S1～SNのマシン負荷情報（すなわち、CPUの使用率や内部メモリの使用率）を接続テーブルに格納して管理する（処理T1.1）。ここで、接続テーブルは、例えば代表ポート部1により作成されて当該代表ポート部1のメモリに保持されており、上述のように各サービスデータ配信部S1～SNのマシン負荷情報を格納する（処理T1.2）。

【0026】また、代表ポート部1では、インターネット3を介してクライアント装置4からのアクセスがあった場合には、例えば接続テーブルに格納された各サービスデータ配信部S1～SNのマシン負荷情報を検出し、当該マシン負荷情報を用いて決定される負荷が最も小さいサービスデータ配信部S1～SNを選択して、選択したサービスデータ配信部S1～SNへ当該クライアント装置4からの表示要求をLAN2を介して送信（転送）することにより、クライアント装置4からのアクセスを振り分ける（処理T1.3）。

【0027】ここで、本例の代表ポート部1が各サービスデータ配信部S1～SNのマシン負荷情報を用いて負荷が最も小さいサービスデータ配信部S1～SNを選択する仕方の具体例を示す。すなわち、本例の代表ポート部1では、例えば常駐形のアプリケーションプログラムを実行することで、上記のような負荷分散管理処理等やアクセス振り分け処理（処理T11～処理T13）を行っており、図3に示すCPU使用率の点数表や図4に示す内部メモリ使用率の点数表を参照して、各サービスデ

ータ配信部 S1～SN の負荷分散評価値を算出する。

【0028】上記図3に示したCPU使用率の点数表は例えば代表ポート部1のメモリに格納されており、この点数表には、複数種類のCPU（本例では、“PentiumI00MH”、“PentiumI33MH”、“PentiumII160MH”、“PentiumII233MH”、“PentiumII400MH”、“PentiumIII600MH”といった6種類のCPU）に関して、それぞれのCPUのベース評価の点数（それぞれ“100”、“133”、“160”、“233”、“400”、“600”）と、CPU使用率が10パーセント（%）以下であるときの乗算係数（全てのCPUに共通で“0.9”）と、CPU使用率が20パーセントであるときの乗算係数（全てのCPUに共通で“0.8”）と、CPU使用率が30パーセントであるときの乗算係数（全てのCPUに共通で“0.7”）と、CPU使用率が40パーセント以上であるときの乗算係数（全てのCPUに共通で“0.6”）とが対応付けられて記憶されている。

【0029】また、上記図4に示した内部メモリ使用率の点数表は例えば代表ポート部1のメモリに格納されており、この点数表には、複数種類の内部メモリ（本例では、“sim64M1n”、“sim64M2n”、“sim64M3n”、“sim128M1n”、“sim128M2n”、“sim128M3n”といった6種類の内部メモリ）に関して、それぞれの内部メモリのベース評価の点数（それぞれ“64*1.0”、“64*0.8”、“64*0.7”、“128*1.0”、“128*0.8”、“128*0.7”）と、内部メモリ使用率が10パーセント（%）以下であるときの乗算係数（全ての内部メモリに共通で“0.9”）と、内部メモリ使用率が30パーセントであるときの乗算係数（全ての内部メモリに共通で“0.5”）と、内部メモリ使用率が70パーセントであるときの乗算係数（全ての内部メモリに共通で“0.3”）と、内部メモリ使用率が80パーセント以上であるときの乗算係数（全ての内部メモリに共通で“0.2”）とが対応付けられて記憶されている。

【0030】具体的には、代表ポート部1では、次のようにして、各サービスデータ配信部S1～SNの負荷分散評価値を算出する。すなわち、代表ポート部1では、例えば各サービスデータ配信部S1～SNに備えられたCPUの種類や内部メモリの種類をメモリに記憶しており、各サービスデータ配信部S1～SNに関して、当該サービスデータ配信部S1～SNに備えられたCPUのベース評価の点数にCPU使用率に応じた乗算係数を乗算した乗算結果と、当該サービスデータ配信部S1～SNに備えられた内部メモリのベース評価の点数に内部メモリ使用率に応じた乗算係数を乗算した乗算結果とを総和した値を算出し、このようにして算出した総和値を各サービスデータ配信部S1～SNの負荷分散評価値とする。

【0031】一例として、負荷分散評価値の算出対象となるサービスデータ配信部S1～SNに備えられたCPUの種類が“PentiumI33MH”であって当該CPUの使用率が20パーセントであるとともに、当該サービスデータ配信部S1～SNに備えられた内部メモリの種類が“sim64M2n”であって当該内部メモリの使用率が50パーセントである場合には、負荷分散評価値は132（＝133*0.8+64*0.8*0.5）と算出される。

【0032】なお、CPU使用率が上記図3の点数表で示した使用率以外の値となるときや、内部メモリ使用率が上記図4の点数表で示した使用率以外の値となるときには、例えば最も近い使用率について設定されている乗算係数が負荷分散評価値の算出に用いられる。また、上記図3に示したCPU使用率の点数表や上記図4に示した内部メモリ使用率の点数表としては、必ずしも本例で示したものに限られずに種々であってもよく、例えば更に細かい使用率（パーセント）に区切って乗算係数を設定したものをを用いることもできる。また、これらの点数表は、例えばシステムの使用状況等に応じて、代表ポート部1により書き換え可能な構成であってもよい。

【0033】また、上記のようにして算出される負荷分散評価値は、例えば各サービスデータ配信部S1～SNの識別情報（マシンID）と対応付けられて、代表ポート部1のメモリに保持される負荷分散検索テーブルに格納される。そして、代表ポート部1では、例えば全てのサービスデータ配信部S1～SNの負荷分散評価値が負荷分散検索テーブルに格納された状態において、負荷分散評価値が最も大きいサービスデータ配信部S1～SNを検索し、当該サービスデータ配信部S1～SNを負荷が最も小さいサービスデータ配信部として選択する。このようにして負荷が最も小さいサービスデータ配信部S1～SNを選択すると、上述したように、代表ポート部1では、クライアント装置4からの表示要求をLAN2を介して当該サービスデータ配信部S1～SNへ送信する。

【0034】次いで、各サービスデータ配信部S1～SNが代表ポート部1を介して受信するクライアント装置4からの表示要求に応じて、当該クライアント装置4へ表示データを送信する処理の具体例を示す。すなわち、各サービスデータ配信部S1～SNでは、例えば代表ポート部1を介して送信されてくるクライアント装置4からの表示要求をLAN2を介して受信すると、当該表示要求を解析して当該表示要求に応じたサービスデータを表示させるためのサービスデータ表示画面のデータ（表示データ）を作成し、作成した表示データをLAN2やインターネット3を介して当該クライアント装置4へ送信する（処理T4）。

【0035】また、上記のようにしてサービスデータ配信部S1～SNが表示データをクライアント装置4へ送

信すると当該送信処理の実行によりCPUの使用率や内部メモリの使用率が変化するが、本例の各サービスデータ配信部S1～SNでは、このような変化があった場合においても、自己のマシン負荷情報を取得して当該マシン負荷情報を更新するとともに、更新後のマシン負荷情報を代表ポート部1に通知する。そして、本例では、このようなマシン負荷情報のリアルタイムなフィードバックにより代表ポート部1で管理されるマシン負荷情報が更新され、このような更新情報が次のクライアント装置4からの表示要求の振り分け処理に有効に反映させられる。

【0036】以上のようにして、本例のデータ配信システムでは、クライアント装置4からインターネット3を介して代表ポート部1へ送信される表示要求に応じて、代表ポート部1が複数のサービスデータ配信部S1～SNの中から選択したサービスデータ配信部S1～SNにより表示データをLAN2やインターネット3を介してクライアント装置4へ送信させる。

【0037】これに際して、本例のデータ配信システムでは、各サービスデータ配信部S1～SNの性能に直結するCPUの使用率や内部メモリの使用率に関する情報が代表ポート部1により常に管理されて、これらの情報を用いて決定される負荷分散評価値に基づいてデータ送信を行わせるサービスデータ配信部S1～SNが選択されるため、各サービスデータ配信部S1～SNの実際の負荷と一致する正確な負荷に基づいて多数のクライアント装置4からの表示要求を各サービスデータ配信部S1～SNに振り分けることができ、これにより、サービスデータ配信部S1～SNの負荷を例えば均等に自動的に分散させることができる。

【0038】また、本例のデータ配信システムでは、各サービスデータ配信部S1～SNの性能に直結するCPUの使用率と内部メモリの使用率との両方を組合せて決定した負荷分散評価値に基づいて負荷が小さいサービスデータ配信部S1～SNが選択されるため、当該選択に用いられる負荷（本例では、負荷分散評価値）が正確であり、負荷分散の精度を高めることができる。また、本例のようなデータ配信システムは、特に、例えばあまり性能が高くないコンピュータから代表ポート部が構成されるような場合や、例えば多数のサービスデータ配信部を設けてシステム性能を向上させるような場合において、大きな効果を発揮することができる。

【0039】ここで、本例の各サービスデータ配信部S1～SNでは、当該サービスデータ配信部S1～SNに備えられたCPUが内部メモリを用いて行う送信処理により表示データをネットワーク（本例では、LAN2やインターネット3）を介してクライアント装置4へ送信する機能により、本発明に言う送信手段が構成されている。

【0040】また、本例の各サービスデータ配信部S1

～SNでは、上記のような送信処理を行うCPUの使用率や当該CPUにより用いられる内部メモリの使用率に関する情報（マシン負荷情報）を所定のタイミングで代表ポート部1に通知する機能により、本発明に言う通知手段が構成されている。

【0041】なお、本例の各サービスデータ配信部S1～SNでは、好ましい態様として、所定の期間毎にマシン負荷情報の通知を行うとともに当該マシン負荷情報の更新があったときにも通知を行う構成としたが、通知を行う所定のタイミングとしては特に限定はなく、例えばCPUの使用率や内部メモリの使用率が所定の値に達したときに通知を行うようにすることもできる。

【0042】また、本例では、好ましい態様として、各サービスデータ配信部S1～SNが自発的にマシン負荷情報を代表ポート部1に通知する構成とすることにより、代表ポート部1と各サービスデータ配信部S1～SNとの間の通信量を少なくしたが、例えば代表ポート部1が各サービスデータ配信部S1～SNに対してマシン負荷情報の通知要求を発する一方、各サービスデータ配信部S1～SNが当該通知要求に応じてマシン負荷情報を代表ポート部1に通知する構成とすることも可能である。

【0043】また、本例の代表ポート部1では、接続テーブル等を用いて各サービスデータ配信部S1～SNから通知されるCPUの使用率や内部メモリの使用率を管理する機能により、本発明に言う管理手段が構成されている。また、本例の代表ポート部1では、クライアント装置4から送信される表示要求に応じて、CPUの使用率や内部メモリの使用率を用いて決定される負荷が小さいサービスデータ配信部S1～SNにより表示データをクライアント装置4へ送信させる機能により、本発明に言う送信制御手段が構成されている。

【0044】なお、本例では、好ましい態様として、負荷が最も小さいサービスデータ配信部S1～SNにより表示データを送信させたが、例えば最小でなくとも、比較的負荷が小さいサービスデータ配信部S1～SNにより送信を行わせることも可能である。一例として、所定の閾値を設けて、負荷が当該閾値以下であるサービスデータ配信部S1～SN（例えば複数ある場合にはいずれか）により送信を行わせる構成とすることも可能である。

【0045】また、本例の各サービスデータ配信部S1～SNはクライアント装置4からの表示要求を受信したことに応じて当該クライアント装置4へ表示データを送信する構成であるため、本例の代表ポート部1では、クライアント装置4からの表示要求を例えばその送信先アドレスを変更してサービスデータ配信部S1～SNへ送信（転送）することをもって、当該サービスデータ配信部S1～SNにより表示データを送信させている。

【0046】ここで、本発明に係るデータ配信システム

や管理装置やサーバ装置の構成としては、必ずしも以上に示したものに限られず、種々なものであってもよく、また、クライアント装置の構成としても同様に種々なものであってもよい。一例として、データ配信システムに設けられるサーバ装置の数としては、特に限定はなく、複数であれば任意であってよい。

【0047】また、本例では、クライアント装置4からの表示要求に応じてサービスデータ配信部S1～SNが表示データを送信する場合を示したが、本発明では、クライアント装置からの要求としてはどのような要求（例えばテキストデータの要求や音声データの要求等）であってよく、また、各サーバ装置から送信されるデータとしても当該要求に応じて種々なものであってもよい。

【0048】また、本例では、各サービスデータ配信部S1～SNから送信される表示データが代表ポート部1を介さずにクライアント装置4へ伝送される構成を示したが、本発明では、例えば各サーバ装置から送信されるデータが管理装置を介してクライアント装置へ伝送される構成とすることも可能である。

【0049】また、本例では、各サービスデータ配信部S1～SNが自己のCPUの使用率や自己の内部メモリの使用率を管理する一方、代表ポート部1がこれら複数のサービスデータ配信部S1～SNのCPU使用率や内部メモリ使用率を管理する構成としたが、本発明では、必ずしもCPUの使用率や内部メモリの使用率がそのままの値で管理等されなくともよく、例えばCPUの使用率や内部メモリの使用率が他の形（例えば使用率の値に所定の演算を施して得られる値や、使用率の値をその大きさに応じてA、B、C、D等の段階へ変換したもの等）で管理等されてもよく、本発明はこのような態様をも包含するものである。

【0050】また、本例では、CPUの使用率と内部メモリの使用率を用いて各サービスデータ配信部S1～SNの負荷を管理する構成としたが、本発明では、例えば各サーバ装置のファイルアクセス頻度等といった他の情報をも用いて負荷を管理する構成とすることも可能である。このような構成では、例えば各サーバ装置では自己のCPU使用率や内部メモリ使用率やファイルアクセス頻度等を検出して管理装置に通知し、管理装置では通知されたこれらの情報を用いて各サーバ装置の負荷を決定

する。

【0051】また、本発明に係る管理装置や各サーバ装置では、例えばCPUや内部メモリ等を備えたハードウェア資源においてCPUがROMに格納された制御プログラムを実行することにより各種の処理を実行するが、本発明は、このような制御プログラムを格納したフロッピーディスクやCD-ROM等のコンピュータにより読み取り可能な記録媒体として把握することもでき、当該制御プログラムを記録媒体からコンピュータに入力してCPUに実行させることにより、本発明に係る処理を遂行させることができる。

【0052】

【発明の効果】以上説明したように、本発明に係るデータ配信システムによると、クライアント装置から送信される要求に応じて管理装置が複数のサーバ装置の中から選択したサーバ装置によりデータをクライアント装置へ送信させるに際して、各サーバ装置が送信処理を行うCPUの使用率や内部メモリの使用率を所定のタイミングで管理装置に通知する一方、管理装置が各サーバ装置から通知されるCPUの使用率や内部メモリの使用率を管理して、CPUの使用率及び内部メモリの使用率を用いて決定される負荷が小さいサーバ装置によりデータをクライアント装置へ送信させるようにしたため、各サーバ装置の実際の負荷と一致する正確な負荷に基づいてサーバ装置の負荷を分散させることができ、効率的な負荷分散を実現することができる。

【図面の簡単な説明】

【図1】本発明の一実施例に係るデータ配信システムの一例を示す図である。

【図2】データ配信システムにより行われる処理の具体例を説明するための図である。

【図3】CPU使用率の点数表の一例を示す図である。

【図4】内部メモリ使用率の点数表の一例を示す図である。

【図5】従来例に係るデータ配信システムの一例を示す図である。

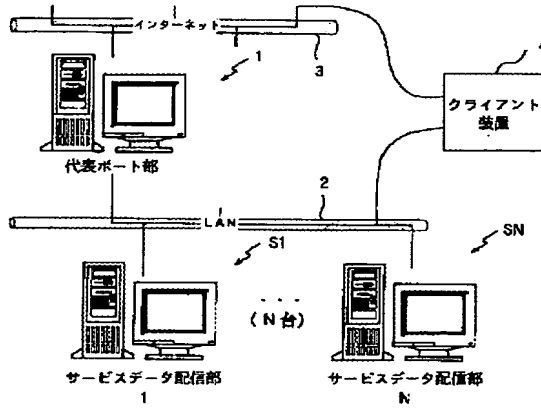
【符号の説明】

1・・・代表ポート部、 S1～SN・・・サービスデータ配信部、 2・・・LAN、 3・・・インターネット、 4・・・クライアント装置、

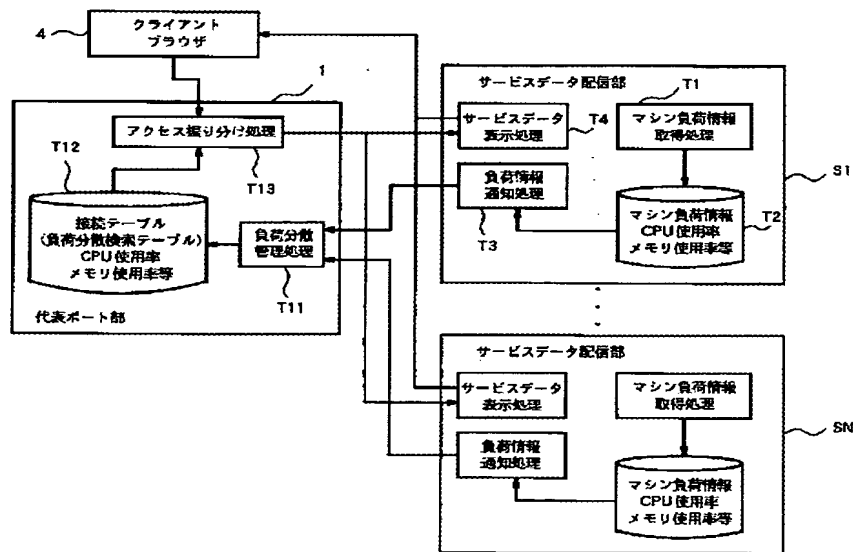
【図3】

項番	CPU	ベース評価	CPU使用率10%以下	CPU使用率20%	CPU使用率30%	CPU使用率40%以上
1	Pentium100MH	100	0.8	0.8	0.7	0.6
2	Pentium133MH	133	0.9	0.8	0.7	0.6
3	Pentium II 160MH	160	0.8	0.8	0.7	0.6
4	Pentium II 233MH	233	0.9	0.8	0.7	0.6
5	Pentium II 400MH	400	0.9	0.8	0.7	0.6
6	Pentium III 600MH	600	0.9	0.8	0.7	0.6

【図1】



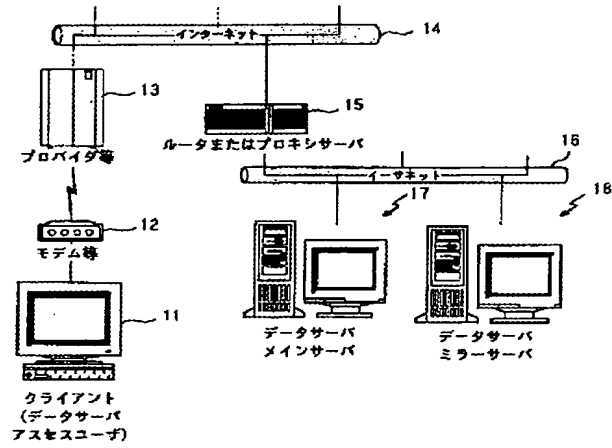
【図2】



【図4】

項目	メモリ	ベース評価	メモリ使用率10%以下	メモリ使用率30%	メモリ使用率50%	メモリ使用率70%	メモリ使用率80%以上
1	simB4M1n	64 * 1.0	0.9	0.7	0.5	0.3	0.2
2	simB4M2n	64 * 0.8	0.9	0.7	0.5	0.3	0.2
3	simB4M3n	64 * 0.7	0.9	0.7	0.5	0.3	0.2
4	sim128M1n	128 * 1.0	0.9	0.7	0.5	0.3	0.2
5	sim128M2n	128 * 0.8	0.9	0.7	0.5	0.3	0.2
6	sim128M3n	128 * 0.7	0.9	0.7	0.5	0.3	0.2

【図5】



フロントページの続き

(51) Int. Cl. ⁷

H04L 12/40

識別記号

F I

テーマコード* (参考)

THIS PAGE BLANK (USPTO)